PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, 15/11, C07K 14/705, C12Q 1/68, C12N 5/10, A61K 48/00, 38/17, G01N 33/50

A2

(11) International Publication Number: **WO 96/39511**

(43) International Publication Date: 12 December 1996 (12.12.96)

(21) International Application Number: PCT/US96/08596

(22) International Filing Date: 3 June 1996 (03.06.96)

(30) Priority Data:
08/462,355 5 June 1995 (05.06.95) US

(71) Applicant: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

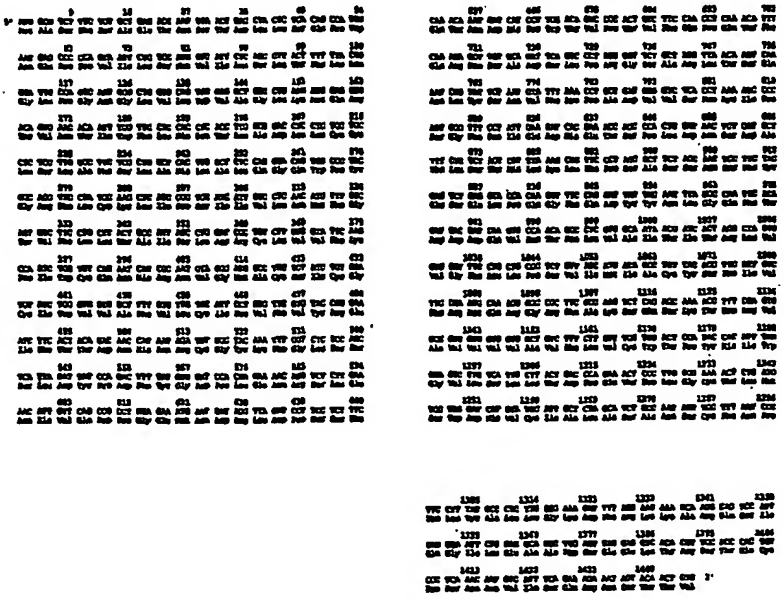
(72) Inventors: COLEMAN, Roger; 260 Mariposa, No. 2, Mountain View, CA 94041 (US). AU-YOUNG, Janice; 1419 Kains Avenue, Berkeley, CA 94702 (US). BANDMAN, Olga; 2309 Rock Street, Mountain View, CA 94043 (US). SEILHAMER, Jeffrey, J.; 12555 La Cresta, Los Altos Hills, CA 94022 (US).

(74) Agent: GLAISTER, Debra, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: A C5a-LIKE SEVEN TRANSMEMBRANE RECEPTOR



(57) Abstract

The present invention provides nucleotide and amino acid sequences that identify and encode a novel C5a-like seven transmembrane receptor (CALR) expressed in human mast cells. The present invention also provides for antisense molecules to the nucleotide sequences which encode CALR, expression vectors for the production of purified CALR, antibodies capable of binding specifically to CALR, hybridization probes or oligonucleotides for the detection of CALR-encoding nucleotide sequences, diagnostic tests based on CALR-encoding nucleic acid molecules, genetically engineered host cells for the expression of CALR, and antagonists and inhibitors of the polypeptide CALR.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

A C5a-LIKE SEVEN TRANSMEMBRANE RECEPTOR

TECHNICAL FIELD

The present invention is in the field of molecular biology; more particularly, the present invention describes the nucleic acid and amino acid sequences of a C5a-like seven transmembrane receptor.

BACKGROUND ART

Complement, which is produced in the liver and circulates in the blood and extracellular fluid, stimulates cells and antibodies to fight infections. Complement 5 (C5) is proteolytically cleaved to produce C5a and C5b whenever the complement system is activated. C5a is one of 13 plasma proteins responsible for clearing foreign particles and organisms from the blood. In addition, human C5a, a 74 amino acid peptide, functions as a chemoattractant for immune system cells.

The C5a receptor is a G-protein coupled seven transmembrane receptor (T7G) which is present on neutrophils, macrophages, and mast cells and is believed to couple with a G_q-/G_{11}-protein to activate the phosphoinositol signaling pathway. The receptor contains 350 amino acids and is glycosylated at Asn^5 to produce a protein of 52-55 kDa. A disulfide bond links Cys^{109} in the first external loop with Cys^{188} in the second external loop. The C5a receptor has been cloned (Boulay et al (1991) Biochem 30:2993-99; Gerard (1991) Nature 349:614-17; and Gerard et al (1992) J Immunol 149:2600-06). Six Asp residues in the N-terminus of the C5a receptor are thought to bind to the Arg and Lys residues in the C5a ligand. With its affinity for peptide ligands and its short third intracellular loop, the C5a receptor most closely resembles the neurokinin T7G receptors.

The T7Gs characteristically contain seven hydrophobic domains which span the plasma membrane and form a bundle of antiparallel α helices. These transmembrane segments are designated by roman numerals I-VII and account for structural and functional features of the receptor. In most cases, the bundle forms a binding pocket; however, when the binding site must accommodate more bulky molecules, the extracellular N-terminal segment or one or more of the three extracellular loops participate in binding (Watson S and Arkinstall S (1994) The G-Protein Linked Receptor Facts Book, Academic Press, San Diego CA) and in subsequent induction of conformational change in intracellular portions of the receptor. The activated receptor in turn, interacts with an intracellular G-protein complex which mediates further intracellular signaling activities, generally the production of second messengers such as cyclic AMP (cAMP), phospholipase C, inositol triphosphate, or ion channel proteins.

Neurokinin-type receptors include tachykinin (TK), formyl peptide (fMLP), GnRH, and prostaglandin E receptors. They are large ligands, mostly peptides, which do not fit the binding pocket of T7G. The N-termini and first extracellular loops have a common tachykinin motif recognition site while the second and third extracellular loops bind to hormone-specific sequences which differ among the receptors. The C-terminus which is common to all isoforms binds to transmembrane helices and activates the receptors. The third intracellular loop is quite short in this group; and in fMLP, it is only 15 amino acids in length. Many of these receptors have short C-termini, and GnRH completely lacks the C-terminal domain (Bolander FF (1994) Molecular Endocrinology, Academic Press, San Diego CA).

The identification of novel C5a-like receptor provides the opportunity to diagnose or intervene in those pathologic or physiologic conditions in which such receptors are expressed or otherwise actively involved.

DISCLOSURE OF THE INVENTION

The subject invention provides a unique nucleotide sequence which encodes a novel human C5a-like receptor homolog, herein designated CALR. The cDNA, herein designated calr, was identified and cloned using Incyte Clone No. 8118 from a human mast cell cDNA library.

The invention also relates to the use of the nucleotide sequence or amino acid sequence of CALR or its variants in the diagnosis or treatemtn of conditions or diseases associated with CALR expression or signal transduction activity. Aspects of the invention include the antisense DNA of calr; cloning or expression vectors containing calr; host cells or organisms transformed with expression vectors containing calr; and a method for the production and recovery of purified CALR protein from host cells. Purified CALR can be used to produce antibodies, antagonists or inhibitors for diagnostic or therapeutic use.

BRIEF DESCRIPTION OF DRAWINGS

Figures 1A, 1B and 1C show the alignment between the nucleotide (SEQ ID NO:1) and amino acid (SEQ ID NO :2) sequences for CALR. The oligomers used to extend the partial nucleotide sequence to full length were XLR = GAAAGACAGCCACCACCACCACG and XLF= AGAAAGCAAGGCAGTCCATTCAGG.

Figure 2 displays the alignment of human CALR with CFOMC5AM, C5A anaphylatoxin receptor from dog; boxed residues are identical.

MODES FOR CARRYING OUT THE INVENTION

As used herein, CALR refers to a C5a-like receptor homolog in naturally occurring or

synthetic form and active fragments thereof, which have the sequence shown in SEQ ID NO:2. In one embodiment, the polypeptide (designated by the upper case, CALR) is encoded by mRNAs transcribed from the cDNA (designated by the lower case, calr) of SEQ ID NO:1.

"Active" refers to those forms of CALR which retain the biologic and/or immunologic activities of any naturally occurring CALR.

"Naturally occurring CALR" refers to CALRs produced by human cells that have not been genetically engineered and specifically contemplates various CALRs arising from post-translational modifications of the polypeptide including but not limited to acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

"Derivative" refers to CALRs chemically modified by such techniques as ubiquitination, labeling (eg, with radionuclides, various enzymes, etc.), pegylation (derivatization with polyethylene glycol), and insertion or substitution by chemical synthesis of amino acids such as ornithine which do not normally occur in human proteins.

"Recombinant variant" refers to any polypeptide having the activity of the CALR protein and differing from naturally occurring CALRs by amino acid insertions, deletions, and substitutions created using recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, such as normal signal transduction, may be found by comparing the sequence of the particular CALR with that of homologous peptides and minimizing the number of amino acid sequence changes made in highly conserved regions.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine, ie, conservative replacements. "Insertions" or "deletions" are typically in the range of about 1 to 5 amino acids. The variation allowed may be experimentally determined by producing the peptide synthetically or by systematically making insertions, deletions, or substitutions of nucleotides in a calr molecule using recombinant DNA techniques and assaying the expressed, recombinant variants for activity.

Where desired, a "signal or leader sequence" can direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous sources by recombinant DNA techniques.

A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, often at least about 7 amino acids, typically at least about 9 to 13 amino acids, and, in various embodiments, at least about 17 or more amino acids. To be active, any CALR peptide must have sufficient length to display biologic and/or immunologic activity.

An "oligonucleotide" or polynucleotide "fragment", "portion", "probe" or "segment" is a stretch of nucleotide residues which is long enough to use in polymerase chain reaction (PCR) or various hybridization procedures. Oligonucleotides are prepared based on the cDNA sequence which encodes CALR provided by the present invention and are used to amplify, or simply reveal the presence of, related RNA or DNA molecules. Oligonucleotides comprise portions of the DNA sequence having at least about 10 nucleotides and as many as about 35 nucleotides, preferably about 25 nucleotides. Nucleic acid probes comprise portions of calr sequence having fewer nucleotides than about 6 kb, preferably fewer than about 1 kb.

After appropriate testing to eliminate false positives, both oligonucleotides and nucleic acid probes may be used to determine whether mRNAs encoding CALR are present in a cell or tissue or to isolate similar natural nucleic acid sequences from chromosomal DNA as described by Walsh PS et al (1992, PCR Methods Appl 1:241-50).

Probes may be derived from naturally occurring or recombinant single- or double-stranded nucleic acids or be chemically synthesized. They may be labeled by nick translation, Klenow fill-in reaction, PCR or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook J et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor NY; or Ausubel FM et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both incorporated herein by reference.

Recombinant variants encoding T7Gs may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce specific restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or to increase expression in a particular prokaryotic or eukaryotic system. Codon usage-specific mutations may also be introduced or chimeras containing the domains of related peptides added to test or modify the properties of any part of the polypeptide, particularly to change ligand-binding affinities, interchain affinities, or degradation/turnover rate.

The present invention provides a unique nucleotide sequence identifying a novel C5a-like receptor which was first identified in human mast cells. The sequence for calr is shown in SEQ ID NO:1 and is homologous to the GenBank sequence, CFCOMC5AM for canine C5a anaphylatoxin receptor. Incyte 8118 has 45% amino acid identity with the C5a receptor and differs from it in having only three carboxylate residues in the N-terminus, two of which are Glu rather than Asp. In addition, the N-terminus of Incyte 8118 is shorter than that of the published C5a receptor and would be expected to have different binding specificity.

Because CALR is expressed in cells active in immunity, the nucleic acid (calr), polypeptide (CALR) and antibodies to CALR are useful in investigations of and interventions in

the normal and abnormal physiologic and pathologic processes associated with the mast cell's role in immunity. Therefore, an assay for upregulated expression of CALR can accelerate diagnosis and proper treatment of conditions caused by abnormal signal transduction events due to anaphylactic or hypersensitive responses, systemic and local infections, traumatic and other tissue damage, hereditary or environmental diseases associated with hypertension, carcinomas, and other physiologic or pathologic problems.

The nucleotide sequence encoding CALR (or its complement) has numerous other applications in techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes for Southern or northern analysis, use as oligomers for PCR, use for chromosomal and gene mapping, use in the recombinant production of CALR, use in generation of antisense DNA or RNA, their chemical analogs and the like, and use in production of chimeric molecules for selecting agonists, inhibitors or antagonists for design of domain-specific therapeutic molecules. Uses of the nucleotides encoding CALR disclosed herein are exemplary of known techniques and are not intended to limit their use in any technique known to a person of ordinary skill in the art. Furthermore, the nucleotide sequences disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, eg, the triplet genetic code, specific base pair interactions, etc.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of CALR-encoding nucleotide sequences, some bearing minimal homology to the nucleotide sequence of any known and naturally occurring gene may be produced. The invention has specifically contemplated each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring CALR, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CALR and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring CALR gene under stringent conditions, it may be advantageous to produce nucleotide sequences encoding CALR or its derivatives possessing a substantially different codon usage. Codons can be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CALR and its derivatives without altering the encoded amino acid sequence include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts

produced from the naturally occurring sequence.

The nucleotide sequence encoding CALR may be joined to a variety of other nucleotide sequences by means of well established recombinant DNA techniques (Sambrook J et al, supra). Useful nucleotide sequences for joining to calr include an assortment of cloning vectors such as plasmids, cosmids, lambda phage derivatives, phagemids, and the like that are well known in the art and may be chosen for such characteristics as the size insert they can accommodate, their utility, their fidelity, etc. Other vectors of interest include expression vectors, replication vectors, probe generation vectors, sequencing vectors, YAC and BAC mapping vectors, and the like. In general, these vectors may contain an origin of replication functional in at least one organism, convenient restriction endonuclease sensitive sites, and selectable markers for recovering transformed host cells.

Another aspect of the subject invention is to provide for calr-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences encoding CALR. Such probes may also be used for the detection of CALR-encoding sequences and should preferably contain at least 50% of the nucleotides from any particular domain of interest from this calr encoding sequence. The hybridization probes of the subject invention may be derived from the nucleotide sequence of the SEQ ID NO:1 or from genomic sequence including promoter, enhancer elements and introns of the respective naturally occurring calrs. Hybridization probes may be labeled by a variety of reporter groups, including radionuclides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

PCR, as described in US Patent Nos. 4,683,195 and 4,965,188, provides additional uses for oligonucleotides based upon the nucleotide sequences which encode CALR. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or may be a mixture of both and comprise a discrete nucleotide sequence for diagnostic use or a degenerate pool of possible sequences for identification of closely related T7G sequences.

Full length genes may be cloned from known sequence using a new method, disclosed in Patent Application Serial No 08/487,112, filed June 7, 1995 and hereby incorporated by reference, which employs XL-PCR (Perkin-Elmer, Foster City, CA) to amplify long pieces of DNA. This method was developed to allow a single researcher to process multiple genes (up to 20 or more) at a time and to obtain an extended (possibly full-length) sequence within 6-10 days. It replaces current methods which use labeled probes to screen libraries and allow one researcher to process only about 3-5 genes in 14-40 days.

In the first step, which can be performed in about two days, primers are designed and synthesized based on a known partial sequence. In step 2, which takes about six to eight hours,

the sequence is extended by PCR amplification of a selected library. Steps 3 and 4, which take about one day, are purification of the amplified cDNA and its ligation into an appropriate vector. Step 5, which takes about one day, involves transforming and growing up host bacteria. In step 6, which takes approximately five hours, PCR is used to screen bacterial clones for extended sequence. The final steps, which take about one day, involve the preparation and sequencing of selected clones. If the full length cDNA has not been obtained, the entire procedure is repeated using either the original library or some other preferred library. The preferred library may be one that has been size-selected to include only larger cDNAs or may consist of single or combined commercially available libraries, eg. lung, liver, heart and brain from Gibco/BRL (Gaithersburg MD). The cDNA library may have been prepared with oligo d(T) or random primers. The advantage of using random primed libraries is that they will have more sequences which contain 5' ends of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a complete gene. Obviously, the larger the protein, the less likely it is that the complete gene will be found in a single plasmid.

Other means for producing hybridization probes for T7G DNAs include the cloning of nucleic acid sequences encoding CALR or its derivatives into vectors for the production of mRNA probes. Such vectors are well known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate labeled nucleotides.

It is now possible to produce a DNA sequence, or portions thereof, encoding CALR and/or its derivatives entirely by synthetic chemistry. Such molecules can be inserted into any of the many available vectors using reagents and methods that are known in the art at the time of the filing of this application. Moreover, synthetic chemistry may be used to introduce mutations into the calr sequences or any portion thereof.

The nucleotide sequence can be used to develop an assay to detect activation, inflammation, or disease associated with abnormal levels of CALR expression. The nucleotide sequence can be labeled by methods known in the art and added to a fluid or tissue sample from a patient. After an incubation period sufficient to effect hybridization, the sample is washed with a compatible fluid which contains a visible marker, a dye or other appropriate molecule(s), if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye is significantly elevated (or lowered, as the case may be), the nucleotide sequence has hybridized with the sample, and the assay indicates an abnormal condition such as inflammation or disease.

The nucleotide sequence for calr can be used to construct hybridization probes for mapping the gene. The nucleotide sequence provided herein may be mapped to a chromosome or

specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York City.

Fluorescent in situ hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of calr on a physical chromosomal map and a specific disease (or predisposition to a specific disease) can help delimit the region of DNA associated with that genetic disease. The nucleotide sequence of the subject invention may be used to detect differences in the genetic sequence between normal and carrier or affected individuals.

The nucleotide sequence encoding CALR may be used to produce purified CALR using well known methods of recombinant DNA technology. Among the many publications that teach methods for the expression of genes after they have been isolated is Goeddel (1990) Gene Expression Technology, Methods and Enzymology, Vol 185, Academic Press, San Diego CA. CALR may be expressed in a variety of host cells, either prokaryotic or eukaryotic. Host cells may be from the same species in which calr nucleotide sequences are endogenous or from a different species. Advantages of producing CALR by recombinant DNA technology include obtaining adequate amounts of the protein for purification and the availability of simplified purification procedures.

Cells transformed with DNA encoding CALR may be cultured under conditions suitable for the expression of CALR and recovery of the protein from the cell culture. CALR produced by a recombinant cell may be secreted or may be contained intracellularly depending on the particular genetic construction used. In general, it is more convenient to prepare recombinant proteins in secreted form. Purification steps vary with the production process and the particular protein produced.

Various methods for the isolation of CALR polypeptide may be accomplished by procedures well known in the art. For example, such a polypeptide may be purified by immunoaffinity chromatography by employing the antibodies provided by the present invention. Various other methods of protein purification well known in the art include those described in Deutscher M (1990) Methods in Enzymology, Vol 182, Academic Press, San Diego CA; and in Scopes R (1982) Protein Purification: Principles and Practice, Springer-Verlag, New York City, both incorporated herein by reference.

In addition to recombinant production, fragments of CALR may be produced by direct peptide synthesis using solid-phase techniques (cf Stewart et al (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco CA; Merrifield J (1963) J Am Chem Soc 85:2149-2154). In vitro protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (ABI, Foster City, California) in accordance with the instructions provided by the manufacturer. Various fragments of CALR may be chemically synthesized separately and combined using chemical methods to produce the full length polypeptide.

CALR for antibody induction does not require biological activity; however, the protein must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. They should mimic an exposed structural portion of the amino acid sequence (an epitope) of the protein and may contain the entire amino acid sequence of a small domain of CALR. Short stretches of CALR amino acids may be fused with those of another protein such as keyhole limpet hemocyanin, and antibody produced against the fusion protein.

Antibodies specific for CALR may be produced by inoculation of an appropriate animal with the polypeptide or an antigenic fragment. An antibody is specific for CALR if it is specific for an immunogenic epitope of the polypeptide and binds to at least part of the natural or recombinant protein. Antibody production includes not only the stimulation of an immune response by injection into animals, but also analogous steps in the production of synthetic antibodies or other specific-binding molecules such as the screening of recombinant immunoglobulin libraries (Orlandi R et al (1989) PNAS 86:3833-37, or Huse WD et al (1989) Science 256:1275-81) or the in vitro stimulation of lymphocyte populations. Current technology (Winter G and Milstein C (1991) Nature 349:293-99) provides for a number of highly specific binding reagents based on the principles of antibody formation. These techniques may be adapted to produce molecules specifically binding particular domains of CALR.

An additional embodiment of the subject invention is the use of CALR specific antibodies or the like as bioactive agents to treat abnormal signal transduction events associated with anaphylactic or hypersensitive responses systemic and local infections, traumatic and other tissue damage, hereditary or environmental diseases associated with hypertension, carcinomas, and other physiologic/pathologic problems.

Bioactive compositions comprising agonists, antagonists, or inhibitors of CALR may be administered in a suitable therapeutic dose determined by any of several methodologies including clinical studies on mammalian species to determine maximum tolerable dose and on

normal human subjects to determine safe dosage. Additionally, the bioactive agent may be complexed with a variety of well established compounds or compositions which enhance stability or pharmacological properties such as half-life. It is contemplated that a therapeutic, bioactive composition may be delivered by intravenous infusion into the bloodstream or any other effective means which could be used for treatment.

The examples below are provided to describe the subject invention. These examples are provided by way of illustration and are not included for the purpose of limiting the invention.

INDUSTRIAL APPLICABILITY

I Isolation of mRNA and Construction of the cDNA Library

The CALR sequence of this application was first identified in Incyte Clone 8118 (SEQ ID NO:1) among the sequences comprising the human mast cell library. The cells used to prepare the human mast cell library were obtained from a Mayo Clinic cancer patient. The mast cell cDNA library was prepared by purifying poly-A^+ mRNA and synthesizing double stranded complementary DNA enzymatically. Synthetic adapters were ligated to the blunt-ended cDNAs which were then ligated to the phage lambda-derived Uni-ZAP™ vector (Stratagene, La Jolla CA). This allowed high efficiency unidirectional (sense orientation) lambda library construction and the convenience of a plasmid system with blue/white color selection to detect clones with cDNA insertions.

The quality of the cDNA library was screened using DNA probes, and then, the pBluescript® phagemid (Stratagene) was excised. This phagemid allows the use of a plasmid system for easy insert characterization, sequencing, site-directed mutagenesis, the creation of unidirectional deletions and expression of fusion polypeptides. Subsequently, the custom-constructed library phage particles were infected into E. coli host strain XL1-Blue® (Stratagene). The high transformation efficiency of this bacterial strain increases the probability that the cDNA library will contain rare, under-represented clones. Alternative unidirectional vectors might include, but are not limited to, pcDNAI (Invitrogen, San Diego CA) and pSHlox-1 (Novagen, Madison WI).

II Isolation of cDNA Clones

The phagemid forms of individual cDNA clones were obtained by the in vivo excision process, in which XL1-BLUE was coinfected with an f1 helper phage. Proteins derived from both lambda phage and f1 helper phage initiate new DNA synthesis from defined sequences on the lambda target DNA and create a smaller, single-stranded circular phagemid DNA molecule that includes all DNA sequences of the pBluescript plasmid and the cDNA insert. The phagemid DNA

was released from the cells and purified, then used to reinfect fresh bacterial host cells (SOLR™ Stratagene), where the double-stranded phagemid DNA was produced. Because the phagemid carries the gene for ß-lactamase, the newly transformed bacteria were selected on medium containing ampicillin.

Phagemid DNA was purified using the QIAWELL-8 Plasmid Purification System® from QIAGEN Inc. (Chatsworth CA). This technique provides a rapid and reliable high-throughput method for lysing the bacterial cells and isolating highly purified phagemid DNA. The DNA eluted from the purification resin was suitable for DNA sequencing and other analytical manipulations.

III Sequencing of cDNA Clones

The cDNA inserts from random isolates of the mast cell library were sequenced in part. Methods for DNA sequencing are well known in the art. Conventional enzymatic methods employed DNA polymerase Klenow fragment, SEQUENASE® (US Biochemical Corp, Cleveland OH) or Taq polymerase to extend DNA chains from an oligonucleotide primer annealed to the DNA template of interest. Methods have been developed for the use of both single- and double-stranded templates. The chain termination reaction products were electrophoresced on urea-acrylamide gels and detected either by autoradiography (for radionuclide-labeled precursors) or by fluorescence (for fluorescent-labeled precursors). Recent improvements in mechanized reaction preparation, sequencing and analysis using the fluorescent detection method have permitted expansion in the number of sequences that can be determined per day using machines such as the Catalyst 800 and the Applied Biosystems 377 or 373 DNA sequencers.

IV Homology Searching of cDNA Clones and Deduced Proteins

Each sequence so obtained was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (developed by TRW Inc., Los Angeles CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT™ 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern

Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

Alternatively, BLAST, which stands for Basic Local Alignment Search Tool, is used to search for local sequence alignments (Altschul SF (1993) J Mol Evol 36:290-300; Altschul, SF et al (1990) J Mol Biol 215:403-10). BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. Whereas it is ideal for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

V Identification, Full Length Cloning, Sequencing and Translation

Analysis of INHERIT™ results from randomly picked and sequenced portions of clones from mast cell library identified Incyte 8118 as a homolog of the canine C5a receptor, CFOMC5AM (Perret et al, supra). The cDNA insert comprising Incyte 8118 was fully sequenced and used as the basis for cloning the full length cDNA.

The cDNA of Incyte 8118 was extended to full length using a modified XL-PCR (Perkin Elmer) procedure disclosed in Patent Application Serial No 08/487,112, by Guegler et al. and filed June 7, 1995 and hereby incorporated by reference. Two primers were designed--one to initiate extension in the antisense direction (XLR=GAAAGACAGCCACCACCACCACG) and the other to extend sequence in the sense direction (XLF=AGAAAGCAAGGCAGTCCATTCAGG). The primers allowed the sequence to be extended "outward" from the known sequence. This generated amplicons containing new, unknown nucleotide sequence for the gene of interest. The primers were designed using Oligo 4.0 (National Biosciences Inc, Plymouth MN) to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence

at temperatures of about 68°-72° C. Any stretches of nucleotide sequence which would result in hairpin structures and primer-primer dimerizations were avoided.

The mast cell cDNA library was used as a template, and XLR and XLS primers were used to extend and amplify the 8118 sequence. By following the instructions for the XL-PCR kit and thoroughly mixing the enzyme, reaction mix, etc., high fidelity amplification is obtained. Beginning with 25 pMol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the MJ PTC200 (MJ Research, Watertown MA) and the following parameters:

Step 1	94° C for 60 sec (initial denaturation)
Step 2	94° C for 15 sec
Step 3	65° C for 1 min
Step 4	68° C for 7 min
Step 5	Repeat step 2-4 for 15 additional times
Step 6	94° C for 15 sec
Step 7	65° C for 1 min
Step 8	68° C for 7 min + 15 sec/cycle
Step 9	Repeat step 6-8 for 11 additional times
Step 10	72° C for 8 min
Step 11	4° C (and holding)

At the end of 28 cycles, 50 µl of the reaction mix was removed; and the remaining reaction mix was run for an additional 10 cycles as outlined below:

Step 1	94° C for 15 sec
Step 2	65° C for 1 min
Step 3	68° C for (10 min + 15 sec)/cycle
Step 4	Repeat step 1-3 for 9 additional times
Step 5	72° C for 10 min

A 5-10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration, about 0.6-0.8%, agarose mini-gel to determine which reactions were successful in extending the sequence. Although all extensions potentially contained a full length gene, some of the largest products or bands were selected and cut out of the gel. Further purification involved using a commercial gel extraction method such as QIAQuick™ (QIAGEN). Following recovery of the DNA, Klenow enzyme was used to trim single-stranded, nucleotide overhangs creating blunt ends which facilitated religation and cloning.

After ethanol precipitation, the products were redissolved in 13 µl of ligation buffer. Then, 1µl T_4 DNA ligase (15 units) and 1µl T_4 polynucleotide kinase were added, and the

mixture was incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 µl of appropriate media) were transformed with 3 µl of ligation mixture and cultured in 80 µl of SOC medium (Sambrook J et al, supra). After incubation for one hour at 37° C, the whole transformation mixture was plated on Luria Bertani (LB)-agar (Sambrook J et al, supra) containing carbenicillin at 25 mg/L. The following day, 12 colonies were randomly picked from each plate and cultured in 150 µl of liquid LB/carbenicillin medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 µl of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 µl of each sample was transferred into a PCR array.

For PCR amplification, 15 µl of concentrated PCR mix (1.33X) containing 0.75 units of Taq polymerase, a vector primer and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 times
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid and sequenced.

The cDNA (SEQ ID NO:1) and amino acid (SEQ ID NO:2) sequences for human CALR are shown in Figures 1A-C. Incyte's calr produced a BLAST score of 412 when compared with the C5a receptor sequence and has a probability of 1.8^{-50} that the sequence similarity occurred by chance. This calr homolog also resembles various N-formylpeptide receptors generating BLAST scores ranging from 381 to 363 with probabilities of 7.4^{-46} to 3.2^{-43}. When the translation of CALR was searched against protein databases such as SwissProt and PIR, no exact matches were found. Fig 2 shows the comparison of the human calr sequence with that of the dog C5a receptor, CFOMC5AM.

VI Antisense analysis

Knowledge of the correct, complete cDNA sequence of CALR enables its use as a tool for antisense technology in the investigation of gene function. Oligonucleotides, cDNA or genomic

fragments comprising the antisense strand of calr are used either in vitro or in vivo to inhibit expression of the mRNA. Such technology is now well known in the art, and antisense molecules are designed at various locations along the nucleotide sequences. By treatment of cells or whole test animals with such antisense sequences, the gene of interest can be effectively turned off. Frequently, the function of the gene is ascertained by observing behavior at the intracellular, cellular, tissue or organismal level (eg, lethality, loss of differentiated function, changes in morphology, etc.).

In addition to using sequences constructed to interrupt transcription of a particular open reading frame, modifications of gene expression are obtained by designing antisense sequences to intron regions, promoter/enhancer elements, or even to trans-acting regulatory genes. Similarly, inhibition is achieved using Hogeboom base-pairing methodology, also known as "triple helix" base pairing.

VII Expression of CALR

Expression of calr is accomplished by subcloning the cDNAs into appropriate expression vectors and transfecting the vectors into analogous expression hosts. In this particular case, the cloning vector previously used for the generation of the cDNA library, pBluescript, also provides for direct expression of calr sequences in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met and the subsequent 7 residues of ß-galactosidase. Immediately following these eight residues is an engineered bacteriophage promoter useful for artificial priming and transcription and a number of unique restriction sites, including Eco RI, for cloning.

Induction of the isolated, transfected bacterial strain with IPTG using standard methods produces a fusion protein corresponding to the first seven residues of ß-galactosidase, about 15 residues of "linker", and the peptide encoded within the cDNA. Since cDNA clone inserts are generated by an essentially random process, there is one chance in three that the included cDNA lies in the correct frame for proper translation. If the cDNA is not in the proper reading frame, it is obtained by deletion or insertion of the appropriate number of bases by well known methods including in vitro mutagenesis, digestion with exonuclease III or mung bean nuclease, or the inclusion of an oligonucleotide linker of appropriate length.

Alternatively, the calr cDNA is shuttled into other vectors known to be useful for expression of protein in specific hosts. Oligonucleotide primers containing cloning sites as well as a segment of DNA (about 25 bases) sufficient to hybridize to stretches at both ends of the target cDNA is synthesized chemically by standard methods. These primers are then used to amplify the desired gene segment by PCR. The resulting gene segment is digested with appropriate restriction enzymes under standard conditions and isolated by gel electrophoresis.

Alternately, similar gene segments are produced by digestion of the cDNA with appropriate restriction enzymes. Using appropriate primers, segments of coding sequence from more than one gene are ligated together and cloned in appropriate vectors. It is possible to optimize expression by construction of such chimeric sequences.

Suitable expression hosts for such chimeric molecules include, but are not limited to, mammalian cells such as Chinese Hamster Ovary (CHO) and human 293 cells, insect cells such as Sf9 cells, yeast cells such as Saccharomyces cerevisiae, and bacteria such as E. coli. For each of these cell systems, a useful expression vector includes an origin of replication to allow propagation in bacteria and a selectable marker such as the ß-lactamase antibiotic resistance gene to allow plasmid selection in bacteria. In addition, the vector includes a second selectable marker such as the neomycin phosphotransferase gene to allow selection in transfected eukaryotic host cells. Vectors for use in eukaryotic expression hosts often require RNA processing elements such as 3' polyadenylation sequences if such are not part of the cDNA of interest.

Additionally, the vector contains promoters or enhancers which increase gene expression. Such promoters are host specific and include MMTV, SV40, and metallothionine promoters for CHO cells; trp, lac, tac and T7 promoters for bacterial hosts; and alpha factor, alcohol oxidase and PGH promoters for yeast. Transcription enhancers, such as the rous sarcoma virus enhancer, are used in mammalian host cells. Once homogeneous cultures of recombinant cells are obtained through standard culture methods, large quantities of recombinantly produced CALR are recovered from the conditioned medium and analyzed using chromatographic methods known in the art.

VIII Isolation of Recombinant CALR

CALR is expressed as a chimeric protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequence such as Factor XA or enterokinase (Invitrogen, San Diego CA) between the purification domain and the calr sequence is useful to facilitate expression of CALR.

IX Testing of Chimeric T7Gs

Functional chimeric T7Gs are constructed by combining the extracellular receptive sequences of a new isoform with the transmembrane and intracellular segments of a known isoform. Such chimeric molecules are useful for testing purposes. This concept was

demonstrated by Kobilka et al (1988, Science 240:1310-1316) who created a series of chimeric α2-ß2 adrenergic receptors (AR) by inserting progressively greater amounts of α2-AR transmembrane sequence into ß2-AR. The binding activity of known agonists changed as the molecule shifted from having more α2 than ß2 conformation, and intermediate constructs demonstrated mixed specificity. The specificity for binding antagonists, however, correlated with the source of the domain VII. The importance of T7G domain VII for ligand recognition was also found in chimeras utilizing two yeast α-factor receptors and is significant because the yeast receptors are classified as miscellaneous receptors. Thus, the functional role of specific domains appears to be preserved throughout the T7G family regardless of category.

In parallel fashion, internal segments or cytoplasmic domains from a particular isoform are exchanged with the analogous domains of a known T7G and used to identify the structural determinants responsible for coupling the receptors to trimeric G-proteins (Dohlman et al (1991) Annu Rev Biochem 60:653-88). A chimeric receptor in which domains V, VI, and the intracellular connecting loop from ß2-AR are substituted into α2-AR are shown to bind ligands with α2-AR specificity, but to stimulate adenylate cyclase in the manner of ß2-AR. This demonstrates that for adrenergic-type receptors, G-protein recognition is present in domains V and VI and their connecting loop. The opposite situation was predicted and observed for a chimera in which the V->VI loop from α1-AR replaced the corresponding domain on ß2-AR and the resulting receptor bound ligands with ß2-AR specificity and activated G-protein-mediated phosphatidylinositol turnover in the α1-AR manner. Finally, chimeras constructed from muscarinic receptors also demonstrated that V->VI loop is the major determinant for specificity of G-protein activity (Bolander FF, supra).

Chimeric or modified T7Gs containing substitutions in the extracellular and transmembrane regions have shown that both portions of the receptor determine ligand binding specificity. For example, two Ser residues are conserved in domain V of all adrenergic and D catecholamine receptors and are necessary for potent agonist activity. These serines are believed to be in the T7G binding site and to form hydrogen bonds with the catechol moiety of the agonists. Similarly, an Asp residue present in domain III of all T7Gs which binds biogenic amines is believed to be in the T7G binding site and to form an ion pair with the ligand amine group.

Functional, cloned T7Gs are expressed in heterologous expression systems and their biological activity assessed (Marullo et al (1988) Proc Natl Acad Sci 85:7551-55; King et al (1990) Science 250:121-23). One heterologous system introduces genes for a mammalian T7G and a mammalian G-protein into yeast cells. The T7G was shown to have appropriate ligand specificity and affinity and trigger appropriate biological activation--growth arrest and

morphological changes--of the yeast cells. Incyte sequences for T7G domains are tested in a similar manner.

X Production of CALR Specific Antibodies

Two approaches are utilized to raise antibodies to CALR, and each approach is useful for generating either polyclonal or monoclonal antibodies. In one approach, denatured protein from reverse phase HPLC separation is obtained in quantities up to 75 mg. This denatured protein is used to immunize mice or rabbits using standard protocols; about 100 micrograms are adequate for immunization of a mouse, while up to 1 mg might be used to immunize a rabbit. For identifying mouse hybridomas, the denatured protein is radioiodinated and used to screen potential murine B-cell hybridomas for those which produce antibody. This procedure requires only small quantities of protein, such that 20 mg would be sufficient for labeling and screening of several thousand clones.

In a second approach, the amino acid sequence of an appropriate CALR domain, as deduced from translation of the cDNA, is analyzed to determine regions of high immunogenicity. Oligopeptides comprising appropriate hydrophilic regions, as illustrated in Figure 3, are synthesized and used in suitable immunization protocols to raise antibodies. Analysis to select appropriate epitopes is described by Ausubel FM et al (supra). The optimal amino acid sequences for immunization are usually at the C-terminus, the N-terminus and those intervening, hydrophilic regions of the polypeptide which are likely to be exposed to the external environment when the protein is in its natural conformation.

Typically, selected peptides, about 15 residues in length, are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to keyhole limpet hemocyanin (KLH; Sigma, St Louis MO) by reaction with M-maleimidobenzoyl-N- hydroxysuccinimide ester (MBS; cf. Ausubel FM et al, supra). If necessary, a cysteine is introduced at the N-terminus of the peptide to permit coupling to KLH. Rabbits are immunized with the peptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity by binding the peptide to plastic, blocking with 1% bovine serum albumin, reacting with antisera, washing and reacting with labeled (radioactive or fluorescent), affinity purified, specific goat anti-rabbit IgG.

Hybridomas are prepared and screened using standard techniques. Hybridomas of interest are detected by screening with labeled CALR to identify those fusions producing the monoclonal antibody with the desired specificity. In a typical protocol, wells of plates (FAST; Becton-Dickinson, Palo Alto CA) are coated during incubation with affinity purified, specific rabbit-anti-mouse (or suitable anti-species Ig) antibodies at 10 mg/ml. The coated wells are blocked with 1% BSA, washed and incubated with supernatants from hybridomas. After washing

the wells are incubated with labeled CALR at 1 mg/ml. Supernatants with specific antibodies bind more labeled CALR than is detectable in the background. Then clones producing specific antibodies are expanded and subjected to two cycles of cloning at limiting dilution. Cloned hybridomas are injected into pristane-treated mice to produce ascites, and monoclonal antibody is purified from mouse ascitic fluid by affinity chromatography on Protein A. Monoclonal antibodies with affinities of at least 10e8 Me-1, preferably 10e9 to 10e10 or stronger, are typically produced by standard procedures as described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; and in Goding (1986) Monoclonal Antibodies: Principles and Practice, Academic Press, New York NY, both incorporated herein by reference.

XI Diagnostic Test Using CALR Specific Antibodies

Particular CALR antibodies are useful for investigating signal transduction events and the diagnosis of infectious or hereditary conditions which are characterized by differences in the amount or distribution of CALR or downstream products of an active signaling cascade. Since CALR was found in a human mast cell library, it appears to be upregulated in cell types mainly involved in immune protection or defense.

Diagnostic tests for CALR include methods utilizing antibody and a label to detect CALR in human body fluids, membranes, cells, tissues or extracts of such. The polypeptides and antibodies of the present invention are used with or without modification. Frequently, the polypeptides and antibodies are labeled by joining them, either covalently or noncovalently, with a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and have been reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include US Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins are produced as shown in US Patent No. 4,816,567, incorporated herein by reference.

A variety of protocols for measuring soluble or membrane-bound CALR, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CALR is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, DE et al (1983, J Exp Med 158:1211f).

XII Purification of Native CALR Using Specific Antibodies

Native or recombinant CALR is purified by immunoaffinity chromatography using antibodies specific for CALR. In general, an immunoaffinity column is constructed by covalently coupling the anti-CALR antibody to an activated chromatographic resin.

Polyclonal immunoglobulins are prepared from immune sera either by precipitation with ammonium sulfate or by purification on immobilized Protein A (Pharmacia LKB Biotechnology, Piscataway, NJ). Likewise, monoclonal antibodies are prepared from mouse ascites fluid by ammonium sulfate precipitation or chromatography on immobilized Protein A. Partially purified immunoglobulin is covalently attached to a chromatographic resin such as CnBr-activated Sepharose (Pharmacia, Piscataway NJ). The antibody is coupled to the resin, the resin is blocked, and the derivative resin is washed according to the manufacturer's instructions.

Such immunoaffinity columns are utilized in the purification of CALR by preparing a fraction from cells containing CALR in a soluble form. This preparation is derived by solubilization of whole cells or of a subcellular fraction obtained via differential centrifugation (with or without addition of detergent) or by other methods well known in the art. Alternatively, soluble CALR containing a signal sequence is secreted in useful quantity into the medium in which the cells are grown.

A soluble CALR-containing preparation is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CALR (eg, high ionic strength buffers in the presence of detergent). Then, the column is eluted under conditions that disrupt antibody/CALR binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope such as urea or thiocyanate ion), and CALR is collected.

XIII Drug Screening

This invention is particularly useful for screening therapeutic compounds by using CALR or binding fragments thereof in any of a variety of drug screening techniques. The polypeptide or fragment employed in such a test is either free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide, fragment or chimera as described above. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, are used for standard binding assays. The formation of complexes between CALR and the agent being tested are measured. Alternatively, one examines the diminution in complex formation between CALR and a receptor caused by the agent being tested.

Thus, the present invention provides methods of screening for drugs or any other agents which affect signal transduction events. These methods, well known in the art, comprise

contacting such an agent with CALR polypeptide or a fragment thereof and assaying (i) for the presence of a complex between the agent and the CALR polypeptide or fragment, or (ii) for the presence of a complex between the CALR polypeptide or fragment and the cell. In such competitive binding assays, the CALR polypeptide or fragment is typically labeled. After suitable incubation, free CALR polypeptide or fragment is separated from that present in bound form, and the amount of free or uncomplexed label is a measure of the ability of the particular agent to bind to CALR or to interfere with the formation of the CALR and agent complex.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the CALR polypeptides and is described in detail in European Patent Application 84/03564, published on September 13, 1984, incorporated herein by reference. Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with CALR polypeptide and washed. Bound CALR polypeptide is then detected by methods well known in the art. Alternatively, purified CALR is coated directly onto plates for use in the aforementioned drug screening techniques. In addition, non-neutralizing antibodies are used to capture the peptide and immobilize it on the solid support.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding CALR specifically compete with a test compound for binding to CALR polypeptides or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CALR.

XIV Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact, eg, agonists, antagonists, or inhibitors. Any of these examples can be used to fashion drugs which are more active or stable forms of the polypeptide or which enhance or interfere with the function of a polypeptide in vivo (Hodgson J (1991) Bio/Technology 9:19-21, incorporated herein by reference).

In one approach, the three-dimensional structure of a protein of interest, or of a protein-inhibitor complex, is determined by x-ray crystallography, by computer modeling or, most typically, by a combination of the two approaches. Both the shape and charges of the polypeptide must be ascertained to elucidate the structure and to determine active site(s) of the molecule. Less often, useful information regarding the structure of a polypeptide is gained by modeling based on the structure of homologous proteins. In both cases, relevant structural information is used to design efficient inhibitors. Useful examples of rational drug design

include molecules which have improved activity or stability as shown by Braxton S and Wells JA (1992, Biochemistry 31:7796- 7801) or which act as inhibitors, agonists, or antagonists of native peptides as shown by Athauda SB et al (1993 J Biochem 113:742-46), incorporated herein by reference.

It is also possible to isolate a target-specific antibody, selected by functional assay, as described above, and then to solve its crystal structure. This approach, in principle, yields a pharmacore upon which subsequent drug design is based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids is expected to be an analog of the original receptor. The anti-id is then used to identify and isolate peptides from banks of chemically or biologically produced peptides. The isolated peptides then act as the pharmacore.

By virtue of the present invention, sufficient amount of polypeptide are made available to perform such analytical studies as X-ray crystallography. In addition, knowledge of the CALR amino acid sequence provided herein provide guidance to those employing computer modeling techniques in place of or in addition to x-ray crystallography.

XV Identification of Other Members of the Signal Transduction Complex

Purified CALR is a research tool for identification, characterization and purification of interacting G-proteins, phospholipase C, adenylate cyclase, or other signal transduction pathway proteins. Radioactive labels are incorporated into a selected CALR domain by various methods known in the art and used in vitro to capture interacting molecules. A preferred method involves labeling the primary amino groups in CALR with ^{125}I Bolton-Hunter reagent (Bolton, AE and Hunter, WM (1973) Biochem J 133: 529). This reagent has been used to label various molecules without concomitant loss of biological activity (Hebert CA et al (1991) J Biol Chem 266: 18989; McColl S et al (1993) J Immunol 150:4550-4555).

Labeled CALR is useful as a reagent for the purification of molecules with which it interacts. In one embodiment of affinity purification, membrane-bound CALR is covalently coupled to a chromatography column. Cell-free extract derived from mast cells or putative target cells is passed over the column, and molecules with appropriate affinity bind to CALR. The CALR-complex is recovered from the column, dissociated and the recovered molecule is subjected to N-terminal protein sequencing. This amino acid sequence is then used to identify the captured molecule or to design degenerate oligonucleotide probes for cloning the relevant gene from an appropriate DNA library.

In an alternate method, antibodies are raised against CALR, specifically monoclonal antibodies, as described above. The monoclonal antibodies are screened to identify those which

inhibit the binding between ligands and CALR. These monoclonal antibodies are then used therapeutically.

XVI Use and Administration of Antibodies, Inhibitors, or Antagonists

Antibodies, inhibitors, or antagonists of CALR (or other treatments to limit signal transduction, LST), provide different effects when administered therapeutically. LSTs are formulated in a nontoxic, inert, pharmaceutically acceptable aqueous carrier medium preferably at a pH of about 5 to 8, more preferably 6 to 8, although pH varies according to the characteristics of the antibody, inhibitor, or antagonist being formulated and the condition to be treated. Characteristics of LSTs include solubility of the molecule, half-life and antigenicity/immunogenicity; these and other characteristics aid in defining an effective carrier. Native human proteins are preferred as LSTs, but organic or synthetic molecules resulting from drug screens are equally effective in particular situations.

LSTs are delivered by known routes of administration including but not limited to topical creams and gels; transmucosal spray and aerosol; transdermal patch and bandage; injectable, intravenous and lavage formulations; and orally administered liquids and pills particularly formulated to resist stomach acid and enzymes. The particular formulation, exact dosage, and route of administration is determined by the attending physician and varies according to each specific situation.

Such determinations are made by considering multiple variables such as the condition to be treated, the LST to be administered, and the pharmacokinetic profile of the particular LST. Additional factors which are taken into account include disease state (eg, severity) of the patient, age, weight, gender, diet, time and frequency of administration, drug combination, reaction sensitivities, and tolerance/response to therapy. Long acting LST formulations are administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular LST.

Normal dosage amounts vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature. See US Patent Nos. 4,657,760; 5,206,344; or 5,225,212. Those skilled in the art employ different formulations for different LSTs. Administration to cells such as nerve cells necessitates delivery in a manner different from that to other cells such as vascular endothelial cells.

It is contemplated that abnormal signal transduction in those conditions or diseases which trigger mast cell activity cause damage that is treatable with LSTs. Such conditions, particularly anaphylactic or hypersensitive responses, are treated as discussed above. The LST is also used to treat other systemic and local infections, traumatic tissue damage, hereditary or

environmental diseases associated with allergies, hypertension, carcinoma, and other physiologic/pathologic problems associated with abnormal signal transduction events.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF INVENTION: C5a-Like Seven Transmembrane Receptor

(iii) NUMBER OF SEQUENCES: 2

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
(B) STREET: 3330 Hillview Avenue
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: To Be Assigned
(B) FILING DATE: Filed Herewith

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NO: 08/462,355
(B) FILING DATE: 5-JUN-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Luther, Barbara J.
(B) REGISTRATION NUMBER: 33954
(C) REFERENCE/DOCKET NUMBER: PF-0040 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 415-855-0555
(B) TELEFAX: 415-852-0195

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: Mast Cell
(B) CLONE: 8118

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGCGTCTT TCTCTGCTGA GACCAATTCA ACTGACCTAC TCTCACAGCC ATGGAATGAG 60

CCCCCAGTAA TTCTCTCCAT GGTCATTCTC AGCCTTACTT TTTTACTGGG ATTGCCAGGC 120

AATGGGCTGG TGCTGTGGGT GGCTGGCCTG AAGATGCAGC GGACAGTGAA CACAATTTGG 180

TTCCTCCACC TCACCTTGGC GGACCTCCTC TGCTGCCTCT CCTTGGCCTT CTCGCTGGCT 240

CACTTGGCTC TCCAGGGACA GTGGCCCTAC GGCAGGTTCC TATGCAAGCT CATCCCCTCC 300

ATCATTGTCC TCAACATGTT TGGCAGTGTC TTCCTGCTTA CTGCCATTAG CCTGGATCGC 360

TGTCTTGTGG TATTCAAGCC AATCTGGTGT CAGAATCATC GCAATGTAGG GATGGCCTGC 420

TCTATCTGTG GATGTATCTG GGTGGTGGCT TTTGTGTTGT GCATTCCTGT GTTCGTGTAC 480

CGGGAAATCT TCACTACAGA CAACCATAAT AGATGTGGCT ACAAATTTGG TCTCTCCAGC 540

TCATTAGATT ATCCAGACTT TTATGGGGAT CCACTAGAAA ACAGGTCTCT TGAAAACATT 600

GTTCAGCCGC CTGGAGAAAT GAATGATAGG TTAGATCCTT CCTCTTTCCA AACAAATGAT 660

CATCCTTGGA CAGTCCCCAC TGTCTTCCAA CCTCAAACAT TTCAAAGACC TTCTGCAGAT 720

TCACTCCCTA GGGGTTCTGC TAGGTTAACA AGTCAAAATC TGTATTCTAA TGTATTTAAA 780

CCTGCTGATG TGGTCTCACC TAAAATCCCC AGTGGGTTTC CTATTGAAGA TCACGAAACC 840

AGCCCACTGG ATAACTCTGA TGCTTTTCTC TCTACTCATT TAAAGCTGTT CCCTAGCGCT 900

TCTAGCAATT CCTTCTACGA GTCTGAGCTA CCACAAGGTT TCCAGGATTA TTACAATTTA 960

GGCCAATTCA CAGATGACGA TCAAGTGCCA ACACCCCTCG TGGCAATAAC GATCACTAGG 1020

CTAGTGGTGG GTTTCCTGCT GCCCTCTGTT ATCATGATAG CCTGTTACAG CTTCATTGTC 1080

TTCCGAATGC AAAGGGGCCG CTTCGCCAAG TCTCAGAGCA AAACCTTTCG AGTGGCCGTG 1140

GTGGTGGTGG CTGTCTTTCT TGTCTGCTGG ACTCCATACC ACATTTGGGG AGTCCTGTCA 1200

TTGCTTACTG ACCCAGAAAC TCCCTTGGGG AAAACTCTGA TGTCCTGGGA TCATGTATGC 1260

ATTGCTCTAG CATCTGCCAA TAGTTGCTTT AATCCCTTCC TTTATGCCCT CTTGGGGAAA 1320

GATTTTAGGA AGAAAGCAAG GCAGTCCATT CAGGGAATTC TGGAGGCAGC CTTCAGTGAG 1380

GAGCTCACAC GTTCCACCCA CTGTCCCTCA AACAATGTCA TTTCAGAAAG AAATAGTACA 1440

ACTGTG 1446

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 482 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Ser Phe Ser Ala Glu Thr Asn Ser Thr Asp Leu Leu Ser Gln
 1               5                  10                  15

Pro Trp Asn Glu Pro Pro Val Ile Leu Ser Met Val Ile Leu Ser Leu
            20                  25                  30

Thr Phe Leu Leu Gly Leu Pro Gly Asn Gly Leu Val Leu Trp Val Ala
        35                  40                  45

Gly Leu Lys Met Gln Arg Thr Val Asn Thr Ile Trp Phe Leu His Leu
    50                  55                  60

Thr Leu Ala Asp Leu Leu Cys Cys Leu Ser Leu Ala Phe Ser Leu Ala
65                  70                  75                  80

His Leu Ala Leu Gln Gly Gln Trp Pro Tyr Gly Arg Phe Leu Cys Lys
                85                  90                  95

Leu Ile Pro Ser Ile Ile Val Leu Asn Met Phe Gly Ser Val Phe Leu
            100                 105                 110

Leu Thr Ala Ile Ser Leu Asp Arg Cys Leu Val Val Phe Lys Pro Ile
        115                 120                 125

Trp Cys Gln Asn His Arg Asn Val Gly Met Ala Cys Ser Ile Cys Gly
    130                 135                 140

Cys Ile Trp Val Val Ala Phe Val Leu Cys Ile Pro Val Phe Val Tyr
145                 150                 155                 160

Arg Glu Ile Phe Thr Thr Asp Asn His Asn Arg Cys Gly Tyr Lys Phe
                165                 170                 175

Gly Leu Ser Ser Ser Leu Asp Tyr Pro Asp Phe Tyr Gly Asp Pro Leu
            180                 185                 190

Glu Asn Arg Ser Leu Glu Asn Ile Val Gln Pro Pro Gly Glu Met Asn
        195                 200                 205
```

```
Asp Arg Leu Asp Pro Ser Ser Phe Gln Thr Asn Asp His Pro Trp Thr
    210                 215                 220

Val Pro Thr Val Phe Gln Pro Gln Thr Phe Gln Arg Pro Ser Ala Asp
225                 230                 235                 240

Ser Leu Pro Arg Gly Ser Ala Arg Leu Thr Ser Gln Asn Leu Tyr Ser
                245                 250                 255

Asn Val Phe Lys Pro Ala Asp Val Val Ser Pro Lys Ile Pro Ser Gly
            260                 265                 270

Phe Pro Ile Glu Asp His Glu Thr Ser Pro Leu Asp Asn Ser Asp Ala
        275                 280                 285

Phe Leu Ser Thr His Leu Lys Leu Phe Pro Ser Ala Ser Ser Asn Ser
    290                 295                 300

Phe Tyr Glu Ser Glu Leu Pro Gln Gly Phe Gln Asp Tyr Tyr Asn Leu
305                 310                 315                 320

Gly Gln Phe Thr Asp Asp Asp Gln Val Pro Thr Pro Leu Val Ala Ile
                325                 330                 335

Thr Ile Thr Arg Leu Val Val Gly Phe Leu Leu Pro Ser Val Ile Met
            340                 345                 350

Ile Ala Cys Tyr Ser Phe Ile Val Phe Arg Met Gln Arg Gly Arg Phe
        355                 360                 365

Ala Lys Ser Gln Ser Lys Thr Phe Arg Val Ala Val Val Val Val Ala
    370                 375                 380

Val Phe Leu Val Cys Trp Thr Pro Tyr His Ile Trp Gly Val Leu Ser
385                 390                 395                 400

Leu Leu Thr Asp Pro Glu Thr Pro Leu Gly Lys Thr Leu Met Ser Trp
                405                 410                 415

Asp His Val Cys Ile Ala Leu Ala Ser Ala Asn Ser Cys Phe Asn Pro
            420                 425                 430

Phe Leu Tyr Ala Leu Leu Gly Lys Asp Phe Arg Lys Lys Ala Arg Gln
        435                 440                 445

Ser Ile Gln Gly Ile Leu Glu Ala Ala Phe Ser Glu Glu Leu Thr Arg
    450                 455                 460

Ser Thr His Cys Pro Ser Asn Asn Val Ile Ser Glu Arg Asn Ser Thr
465                 470                 475                 480

Thr Val
```

CLAIMS

1. A purified polynucleotide encoding a polypeptide with the amino acid sequence shown in SEQ ID NO:2.

2. The polynucleotide of Claim 1 wherein the nucleic acid sequence comprises SEQ ID NO:1, or its complement.

3. A diagnostic test for conditions or diseases associated with expression of the human C5a-like receptor (calr) in a biological sample comprising the steps of:

a) combining the biological sample with the polynucleotide of Claim 1, or a fragment thereof, under conditions suitable for the formation of hybridization complex; and

b) detecting the hybridization complex, wherein the presence of the complex correlates with expression of the polynucleotide of Claim 1 in the biological sample.

4. An expression vector comprising the polynucleotide of Claim 1.

5. A host cell transformed with the expression vector of Claim 4.

6. A method for producing a polypeptide comprising the amino acid sequence shown in SEQ ID NO:2, the method comprising the steps of:

a) culturing the host cell of Claim 5 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

7. An antisense molecule comprising the nucleic acid sequence complementary to at least a portion of the polynucleotide of Claim 1.

8. A pharmaceutical composition comprising the antisense molecule of Claim 7 and a pharmaceutically acceptable excipient.

9. A method of treating a subject with a condition or disease involving the altered expression of human C5a-like receptor homolog comprising administering an effective amount of the pharmaceutical composition of Claim 8 to the subject.

10. A purified polypeptide comprising the amino acid sequence of SEQ ID NO:2.

11. An agonist of the polypeptide of Claim 10.

12. A pharmaceutical composition comprising the agonist of Claim 11 and a pharmaceutically acceptable excipient.

13. A method of treating a subject with a condition or disease associated with the altered expression of human C5a-like receptor homolog comprising administering an effective amount of the pharmaceutical composition of Claim 12 to the subject.

14. An inhibitor of the polypeptide of Claim 10.

15. A pharmaceutical composition comprising the inhibitor of Claim 14 and a pharmaceutically acceptable excipient.

16. A method of treating a subject with a condition or disease associated with the altered expression of human C5a-like receptor homolog comprising administering an effective amount of the pharmaceutical composition of Claim 15 to the subject.

17. An antibody specific for the purified polypeptide of Claim 10.

18. A diagnostic test for a condition or disease associated with the expression of the C5a-like receptor homolog in a biological sample comprising the steps of:

a) combining the biological sample with the antibody of Claim 17, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the expression of the polypeptide in the biological sample.

19. A pharmaceutical composition comprising the antibody of Claim 17 and a pharmaceutically acceptable excipient.

20. A method of treating a subject with a condition or disease associated with the altered expression of human C5a-like receptor homolog comprising administering an effective amount of the pharmaceutical composition of Claim 19 to the subject.

```
            9          18          27          36          45          54
5' ATG GCG TCT TTC TCT GCT GAG ACC AAT TCA ACT GAC CTA CTC TCA CAG CCA TGG
   Met Ala Ser Phe Ser Ala Glu Thr Asn Ser Thr Asp Leu Leu Ser Gln Pro Trp

           63          72          81          90          99         108
   AAT GAG CCC CCA GTA ATT CTC TCC ATG GTC ATT CTC AGC CTT ACT TTT TTA CTG
   Asn Glu Pro Pro Val Ile Leu Ser Met Val Ile Leu Ser Leu Thr Phe Leu Leu

          117         126         135         144         153         162
   GGA TTG CCA GGC AAT GGG CTG GTG CTG TGG GTG GCT GGC CTG AAG ATG CAG CGG
   Gly Leu Pro Gly Asn Gly Leu Val Leu Trp Val Ala Gly Leu Lys Met Gln Arg

          171         180         189         198         207         216
   ACA GTG AAC ACA ATT TGG TTC CTC CAC CTC ACC TTG GCG GAC CTC CTC TGC TGC
   Thr Val Asn Thr Ile Trp Phe Leu His Leu Thr Leu Ala Asp Leu Leu Cys Cys

          225         234         243         252         261         270
   CTC TCC TTG GCC TTC TCG CTG GCT CAC TTG GCT CTC CAG GGA CAG TGG CCC TAC
   Leu Ser Leu Ala Phe Ser Leu Ala His Leu Ala Leu Gln Gly Gln Trp Pro Tyr

          279         288         297         306         315         324
   GGC AGG TTC CTA TGC AAG CTC ATC CCC TCC ATC ATT GTC CTC AAC ATG TTT GGC
   Gly Arg Phe Leu Cys Lys Leu Ile Pro Ser Ile Ile Val Leu Asn Met Phe Gly

          333         342         351         360         369         378
   AGT GTC TTC CTG CTT ACT GCC ATT AGC CTG GAT CGC TGT CTT GTG GTA TTC AAG
   Ser Val Phe Leu Leu Thr Ala Ile Ser Leu Asp Arg Cys Leu Val Val Phe Lys

          387         396         405         414         423         432
   CCA ATC TGG TGT CAG AAT CAT CGC AAT GTA GGG ATG GCC TGC TCT ATC TGT GGA
   Pro Ile Trp Cys Gln Asn His Arg Asn Val Gly Met Ala Cys Ser Ile Cys Gly

          441         450         459         468         477         486
   TGT ATC TGG GTG GTG GCT TTT GTG TTG TGC ATT CCT GTG TTC GTG TAC CGG GAA
   Cys Ile Trp Val Val Ala Phe Val Leu Cys Ile Pro Val Phe Val Tyr Arg Glu

          495         504         513         522         531         540
   ATC TTC ACT ACA GAC AAC CAT AAT AGA TGT GGC TAC AAA TTT GGT CTC TCC AGC
   Ile Phe Thr Thr Asp Asn His Asn Arg Cys Gly Tyr Lys Phe Gly Leu Ser Ser

          549         558         567         576         585         594
   TCA TTA GAT TAT CCA GAC TTT TAT GGG GAT CCA CTA GAA AAC AGG TCT CTT GAA
   Ser Leu Asp Tyr Pro Asp Phe Tyr Gly Asp Pro Leu Glu Asn Arg Ser Leu Glu

          603         612         621         630         639         648
   AAC ATT GTT CAG CCG CCT GGA GAA ATG AAT GAT AGG TTA GAT CCT TCC TCT TTC
   Asn Ile Val Gln Pro Pro Gly Glu Met Asn Asp Arg Leu Asp Pro Ser Ser Phe
```

FIGURE 1A

```
        657         666         675         684         693         702
CAA ACA AAT GAT CAT CCT TGG ACA GTC CCC ACT GTC TTC CAA CCT CAA ACA TTT
Gln Thr Asn Asp His Pro Trp Thr Val Pro Thr Val Phe Gln Pro Gln Thr Phe

        711         720         729         738         747         756
CAA AGA CCT TCT GCA GAT TCA CTC CCT AGG GGT TCT GCT AGG TTA ACA AGT CAA
Gln Arg Pro Ser Ala Asp Ser Leu Pro Arg Gly Ser Ala Arg Leu Thr Ser Gln

        765         774         783         792         801         810
AAT CTG TAT TCT AAT GTA TTT AAA CCT GCT GAT GTG GTC TCA CCT AAA ATC CCC
Asn Leu Tyr Ser Asn Val Phe Lys Pro Ala Asp Val Val Ser Pro Lys Ile Pro

        819         828         837         846         855         864
AGT GGG TTT CCT ATT GAA GAT CAC GAA ACC AGC CCA CTG GAT AAC TCT GAT GCT
Ser Gly Phe Pro Ile Glu Asp His Glu Thr Ser Pro Leu Asp Asn Ser Asp Ala

        873         882         891         900         909         918
TTT CTC TCT ACT CAT TTA AAG CTG TTC CCT AGC GCT TCT AGC AAT TCC TTC TAC
Phe Leu Ser Thr His Leu Lys Leu Phe Pro Ser Ala Ser Ser Asn Ser Phe Tyr

        927         936         945         954         963         972
GAG TCT GAG CTA CCA CAA GGT TTC CAG GAT TAT TAC AAT TTA GGC CAA TTC ACA
Glu Ser Glu Leu Pro Gln Gly Phe Gln Asp Tyr Tyr Asn Leu Gly Gln Phe Thr

        981         990         999        1008        1017        1026
GAT GAC GAT CAA GTG CCA ACA CCC CTC GTG GCA ATA ACG ATC ACT AGG CTA GTG
Asp Asp Asp Gln Val Pro Thr Pro Leu Val Ala Ile Thr Ile Thr Arg Leu Val

       1035        1044        1053        1062        1071        1080
GTG GGT TTC CTG CTG CCC TCT GTT ATC ATG ATA GCC TGT TAC AGC TTC ATT GTC
Val Gly Phe Leu Leu Pro Ser Val Ile Met Ile Ala Cys Tyr Ser Phe Ile Val

       1089        1098        1107        1116        1125        1134
TTC CGA ATG CAA AGG GGC CGC TTC GCC AAG TCT CAG AGC AAA ACC TTT CGA GTG
Phe Arg Met Gln Arg Gly Arg Phe Ala Lys Ser Gln Ser Lys Thr Phe Arg Val

       1143        1152        1161        1170        1179        1188
GCC GTG GTG GTG GTG GCT GTC TTT CTT GTC TGC TGG ACT CCA TAC CAC ATT TGG
Ala Val Val Val Val Ala Val Phe Leu Val Cys Trp Thr Pro Tyr His Ile Trp

       1197        1206        1215        1224        1233        1242
GGA GTC CTG TCA TTG CTT ACT GAC CCA GAA ACT CCC TTG GGG AAA ACT CTG ATG
Gly Val Leu Ser Leu Leu Thr Asp Pro Glu Thr Pro Leu Gly Lys Thr Leu Met

       1251        1260        1269        1278        1287        1296
TCC TGG GAT CAT GTA TGC ATT GCT CTA GCA TCT GCC AAT AGT TGC TTT AAT CCC
Ser Trp Asp His Val Cys Ile Ala Leu Ala Ser Ala Asn Ser Cys Phe Asn Pro
```

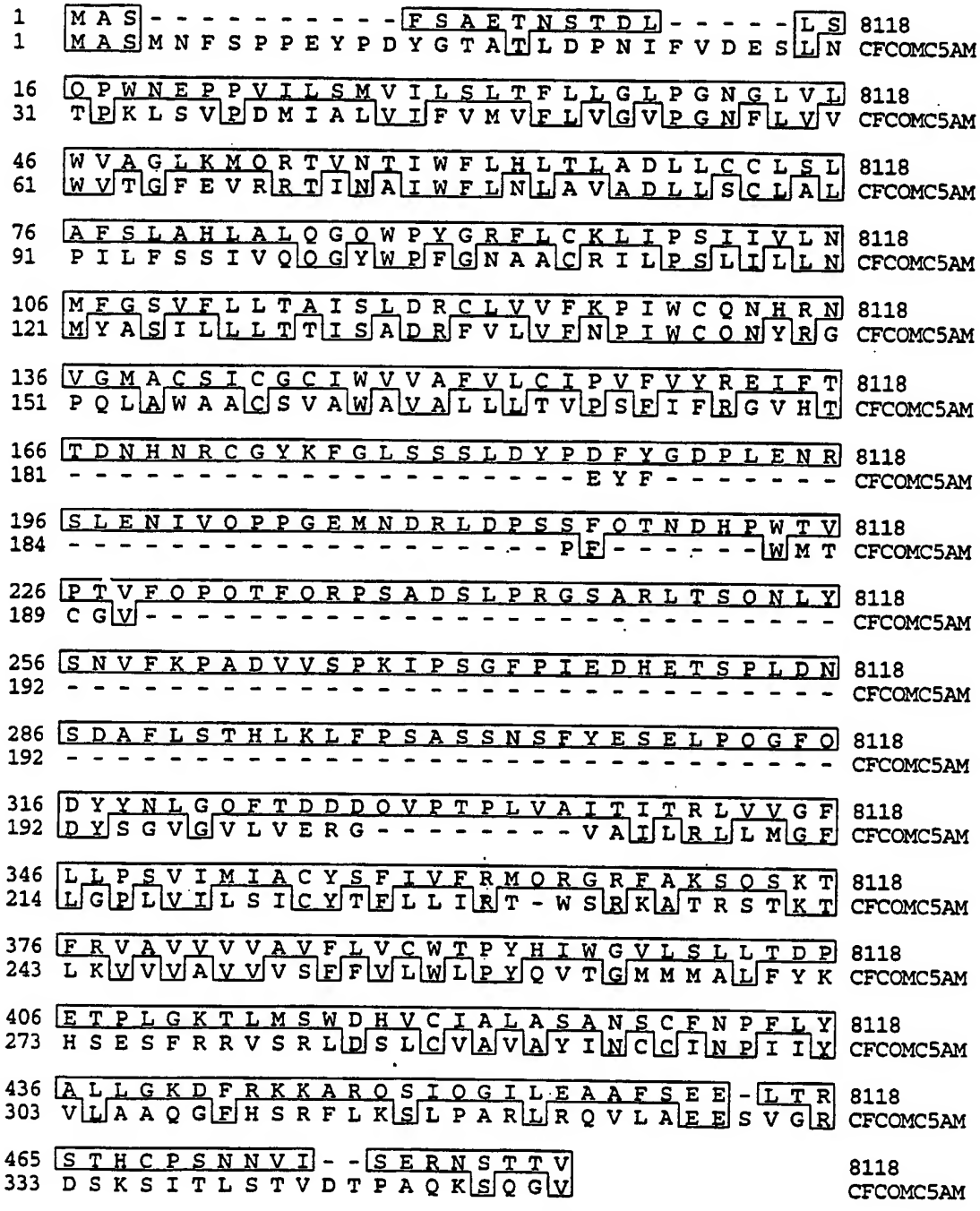
FIGURE 1B

```
      1305        1314        1323        1332        1341        1350
TTC CTT TAT GCC CTC TTG GGG AAA GAT TTT AGG AAG AAA GCA AGG CAG TCC ATT
Phe Leu Tyr Ala Leu Leu Gly Lys Asp Phe Arg Lys Lys Ala Arg Gln Ser Ile

      1359        1368        1377        1386        1395        1404
CAG GGA ATT CTG GAG GCA GCC TTC AGT GAG GAG CTC ACA CGT TCC ACC CAC TGT
Gln Gly Ile Leu Glu Ala Ala Phe Ser Glu Glu Leu Thr Arg Ser Thr His Cys

      1413        1422        1431        1440
CCC TCA AAC AAT GTC ATT TCA GAA AGA AAT AGT ACA ACT GTG  3'
Pro Ser Asn Asn Val Ile Ser Glu Arg Asn Ser Thr Thr Val
```

FIGURE 1C



Decoration 'Decoration #1': Box residues that match 8118 exactly.

FIGURE 2